



International Civil Aviation English Association

ICAEA Workshop 2017, The ICAO LPRs – 10
years on: Progress or Pain?

Apr 24th, 10:35 AM - 10:50 AM

What works and doesn't work in LPR testing: Is there light at the end of the tunnel?

Adrian Enright

ICAEA, adrian.enright@icaea.aero

Gareth Williams

Horizon Flight Academy, UAE, email.garywilliams@yahoo.co.uk

Follow this and additional works at: <https://commons.erau.edu/icaea-workshop>

Scholarly Commons Citation

Enright, Adrian and Williams, Gareth, "What works and doesn't work in LPR testing: Is there light at the end of the tunnel?" (2017). *International Civil Aviation English Association*. 6.
<https://commons.erau.edu/icaea-workshop/2017/monday/6>

This Presentation is brought to you for free and open access by the Conferences at Scholarly Commons. It has been accepted for inclusion in International Civil Aviation English Association by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

What works and what doesn't work in LPR Testing?

Introduction

In April 2017, the authors of this article had the privilege of delivering a presentation and workshop at an ICAEA (International Civil Aviation English Association) symposium in Dubrovnik, Croatia. That symposium, entitled "The ICAO LPRs – 10 years on: Progress or Pain?" challenged the participants to assess the performance of the training and testing of aviation English language since the introduction of ICAO Language Proficiency Requirements in 2003. This article is written as an explanation of, and expansion of, the authors' presentation and workshop.

Our brief was simple; introduce the symposium participants to a few of the basic principles of test development, place them in the context of aviation language proficiency requirements (LPRs) and run a workshop where the participants could share their ideas on what factors influence good test design. Many of the symposium participants were not test experts so our challenge was to stimulate a discussion on the topic of testing of aviation English while simultaneously encouraging everyone to participate and gently informing the audience on testing "best practice."

As far as this article is concerned, it will be helpful to define two concepts before continuing: (1) what is aviation English? And (2) what should be considered a "bad" test in this context?

What is aviation English?

In the context of English language proficiency we may consider aviation English as an application of language for specific purposes. Aeronautical communication between pilots and air traffic controllers is essentially conducted using ICAO (International Civil Aviation Organisation) standard phraseology – ultra-specific, formulaic statements that cover all phases of flight and most situations requiring communication between air traffic controllers and pilots. However, there are situations where plain language (i.e. the spontaneous, creative and non-coded use of a given natural language) is required, and plain language shall be used *"only when standardized phraseology cannot serve an intended transmission"* (ICAO, 2001). Whilst aviation English in general may be considered in the context of other aviation professions, regulated proficiency of English refers specifically to aeronautical communication essential to safe and efficient operations between pilots and air traffic controllers.

What might a "bad test" look like?

Imagine for a moment, a test of English given to pilots that will have a fundamental influence on the test-takers' career. This particular test for pilots, by the way, is real. It is currently used, and it is widespread. Quite literally, thousands of pilots must pass this test (or similar) if they wish to continue their careers in aviation and they must score high marks to have any chance of pursuing "prestige" (i.e. higher paid) positions within their career.

This test is multiple-choice. At no point is the test-taker ever actually required to speak. The first section of the test involves listening to a recording, reading a comprehension question and choosing the correct answer from four options. The second section is also multiple-choice, but is focused on choosing the correct definition of an idiomatic phrase, e.g. "piece of cake", or "on pins and needles." The idiomatic phrases are unique to one particular western country and the test was written in the 1970s (which many of the idioms reflect). The test is also freely available on the Internet.

Hopefully, even if you don't have a formal qualification in test development, you will recognize that

this is what we can safely call a “bad” test. The question is “why?” What is it about this test that makes the hair stand up on the back of our neck?

Upon consideration of the test described above, you might be now asking yourself one or more of the following questions:

- What exactly does this test have to do with the language proficiency requirements of a pilot?
- If pilots pass this multiple-choice written test, could we safely assume those pilots would be good at using idiomatic phrases in their normal speech? And, should we assume that control of idiomatic speech is useful for radiotelephony?
- How confident can we be that people who pass this test will have clear pronunciation? Or, they will be able to produce accurate grammar in their speech?
- Will this test produce consistent results each time it is applied?

In a multiple choice test with four options, where choosing answers at random should produce a score of 25%, how confident can we be that someone who has scored 35% is better at English than someone who has scored 31%? Similarly, how confident can we be that someone who scored 90% is better at English than someone who scored 85%?

In short, can we be sure that a pilot who scores well on this test will be sufficiently proficient in English to communicate safely with air traffic control.

Well... obviously not.

Proficiency language testing of Aviation English is a global requirement of ICAO Contracting States and has a direct impact not only upon aviation safety but also on the careers of pilots and air traffic controllers. It is the responsibility of individual government regulators to choose, or in some cases oversee the development of, aviation English language proficiency tests that they feel meet the needs of their state, in compliance with ICAO requirements. In some states, the regulator recognizes a number of different tests (Huhta, 2009). A combination of sovereignty rights, operational constraints, commercial interests and bureaucratic inertia has led to dozens of different LPR tests being used worldwide. Regrettably, the vast majority of these tests do not go through a rigorous or public quality audit (Alderson, 2010) with the result that many are perceived as being “easy” for candidates with a level 6 guaranteed (i.e. no further testing required).

So how should someone responsible for the implementation of ICAO LPRs in a State (the regulator) approach the problem of testing aviation English?

Concepts of Test Development

It is helpful for regulators to become familiar with a few fundamental factors of test development. Tests can be evaluated using the concepts of (1) validity, (2) reliability and (3) practicality.

1. Validity

The extent to which scores on a test enable inferences to be made about language proficiency which are appropriate, meaningful and useful given the purpose of the test (ICAO, 2010). Validity relates to how confident we can be, that the test has produced relevant results. Or, in other words, can we make useful predictions about the test participants’ language performance based upon the results of

their tests?

1.1 Content Validity

Content validity (Nunan, 2013) is achieved by choosing test tasks that replicate the language tasks that ICAO has determined are essential in aeronautical communication. Therefore, the test tasks should all relate to radiotelephony and, more specifically, to the ICAO defined concept of “plain language.” In short, aviation English tests should focus on the language required by a pilot and an air traffic controller to resolve a situation where standard ICAO phraseology is not sufficient to operate flights safely.

1.2 Criterion-related validity

Criterion-related validity is achieved by comparing the tests results to either future performance, or to a separate widely valued measurement. For example, such an approach was attempted by Dusenbury and Bjerke (2013) when they compared IELTS scores of ab initio pilot cadets to their future likelihood of passing flight training.

1.3 Construct validity

Content validity is a necessary, but not sufficient, aspect of a useful test. Knowledge of language by itself is often not sufficient for effective communication (Munby, 1978). This is where construct validity comes into play. A construct is a proficiency or skill that is derived from the human brain. Consider the earlier described “bad” test. This test lacks content validity because it tests knowledge of idioms, and idioms are not what air traffic controllers and pilots require for efficient communication maintaining safe and expeditious flight. If this test was changed so that it tested examples of “plain language” we might consider it, therefore, to have content validity. But since the test is still a multiple choice test, and we know that pilots need to actually speak to air traffic controllers to do their job, we cannot say that the test would then have construct validity. Choosing A, B, C or D correctly on a worksheet is not a proficiency or skill that is needed for the safe operation of aircraft.

2. Reliability

Reliability is essentially a measurement of how consistent, or stable, the test is (Bachman & Palmer, 2013). In other words, will the test produce the same results on any given iteration of its application? For example, if 1000 pilots were assessed on two different days, the test would be deemed reliable if the same pilots scored the same results on the second day. It is generally considered unrealistic to eliminate all inconsistencies of a test. However, it is the goal of principled test developers to minimize inconsistencies as far as possible (Bachman & Palmer, 2013). This is primarily achieved through test design and proven through analysis of the data generated after significant numbers of iterations of the test. The results from different sections of a test can be compared to each other in order to achieve internal consistency (Brindley, 2011).

3. Practicality

Put simply, a 100% valid and reliable test will often become impractical to implement. 100% validity in particular might well require language testing to be conducted in a simulator and involve spontaneous communications between air traffic controllers and pilots in simulated non-standard flight scenarios. A practical test must be easy, cost effective and time efficient to administer.

For regulators who are attempting to evaluate their options as regards the quality of any particular aviation English test, understanding the concepts of validity, reliability and practicality will help them to make an informed decision. Regulators don't need to be test experts to ask some simple, straightforward questions of test developers, e.g.:

Is this test valid? If so, demonstrate proof.

Is this test reliable? If so, demonstrate proof.

Is this test practical to administer? What might be the costs involved?

The Workshops

Regulators should consider a variety of factors when choosing or developing aviation English tests for accreditation in their state. The purpose of the workshop conducted by this article's authors was to allow participants to share ideas as to what those factors might be.

The workshop was conducted twice, with two separate groups of participants (total 70 persons). Each group was tasked with determining factors that influence LPR test design and development. The findings from each group were similar in the topics identified but when asked to select (in their view) the three most important factors there was a difference:

Group A: Clear construct, reliability and validity.

Group B: Financial resources, feedback and anonymous non-biased rating.

The factors identified provide prima facie evidence of the composition of each group. Group A appears to have zeroed in on classic testing principles and this might reflect their professional background in the field of language testing. Group B, on the other hand, has prioritized more practical considerations and it is possible that the composition of this group leaned towards operational expertise. Such specific conclusions should be made very tentatively. Either way, the workshop demonstrated a vital consideration of aviation English testing: aircrew, regulators and test experts must all be given a voice as these issues are debated.

Group A's primary assertion deserves further examination. They concluded that having a "clear construct" is a vital factor influencing regulators decisions about aviation English tests. Orally, the participants of this group explained that there is no "clear construct" established. This might be described as an overarching issue affecting the entire industry. In layman's terms, although we know that we need to test pilots' and air traffic controllers' ability to conduct efficient communication maintaining safe and expeditious flight, what sub-skills will achieve this? Obviously further research in this vital area will be of benefit to all.

Most likely the differences between the groups were down to the composition of the participants and their different professional backgrounds. We were unable in the time allotted to explore this further. However, all participants agreed that, in many countries, a lack of awareness of a test needing to be valid, reliable and practical is a factor influencing the selection of language proficiency tests that in most cases do not meet the ICAO requirements as laid out in Doc 9835 (Manual of the Implementation of ICAO Language Proficiency Requirements). Regulators need a basic understanding of testing best practices in order to make informed decisions.

Conclusion

If participation and debate are hallmarks of a valid workshop, we might suggest that this one was a success. Participants came away with a deeper understanding of testing best practices, including

basic concepts of validity, reliability and practicality. Participants also articulated a wide variety of factors that influence test design and/or selection. This demonstrated a depth of knowledge among many participants that other non-test experts present were able to benefit from. It was a rare opportunity for regulators and aviation professionals to gain further insight into the testing of aviation English. Furthermore, it also provided a platform for these aviation professionals and regulators to provide their own unique, and vital, perspectives. In the context of aeronautical communications, all stakeholders' points of view must be carefully considered. We would like to take this opportunity to thank ICAEA for organizing the event and to all of our colleagues who contributed towards the success of this workshop.

- Alderson, J. (2010). A Survey of Aviation English Tests. *Language Testing*, 27(1).
- Bachman, L., & Palmer, A. (2013). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: OUP.
- Brindley, G. (2011). Assessment. In D. Nunan (Ed.), *The Cambridge Guide to Teaching English to Speakers of Other Languages*. Cambridge: CUP.
- Dusenbury, M., & Bjerke, E. (2013). Predictive power of English testing: Training international flight students. *The Journal of Aviation/Aerospace Education & Research*, 23(1).
- Huhta, A. (2009). An analysis of the quality of English testing for aviation purposes in Finland. *Australian Review of Applied Linguistics*, 32(3).
- ICAO. (2001). *Annex 10 to the Convention on International Civil Aviation: Aeronautical Telecommunications Volume II: Communication Procedures including those with PANS status (Sixth Edition)*. Montréal: International Civil Aviation Organisation.
- ICAO. (2010). *Manual on the Implementation of ICAO Language Proficiency Requirements (Second Edition)*. (Doc. 9835). Montreal: International Civil Aviation Organisation.
- Munby, J. (1978). *Communicative Syllabus Design*. Cambridge: CUP.
- Nunan, D. (2013). *Task-Based Language Teaching*. Cambridge: CUP.